CF Conformance Checking: CMOR and other Tools

Karl E. Taylor

Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory

Global Organization for Earth System Science Portal (GO-ESSP) Workshop

PCMDI, Livermore, CA

19-21 June 2006

Design challenges for CF conformance verification

- CF fundamentals:
  - → The conventions specify the syntax and vocabulary for climate and forecast metadata.
  - → Most metadata accommodated by CF is ***not*** mandatory, hence conformance can be trivial, but not particularly useful.
- Decisions (tradeoffs) affecting verification software.
  - → General conformance checker can only flag noncompliance of mandatory metadata, so it is unable to encourage “good practice”. Example: CF compliance checker (Rosalyn Hatcher)
  - → Single project checker can provide comprehensive check, but must be modified for each project. Example: IPCC checker

IPCC AR4 approach to data compliance verification:

- Encourage use of CMOR (Climate Model Output Rewriter) which traps/prevents many errors at the “source”.
- Require “sample” files to be produced and sent to PCMDI where
  - → An IPCC project-specific CF checker was written and applied to samples (also available for use by modeling centers themselves).
  - → Some human inspection of the sample data (e.g., of various plots) uncovered many problems, but was tedious and not comprehensive.

Errors were flagged by CMOR which:

- Points out when required metadata are omitted.
- Rejects incorrect metadata (wrong units, inadmissible attribute values, etc.)
- Rejects inconsistent coordinate dimensions passed by user to CMOR.
- Rejects non-monotonic coordinate values or inconsistent boundary values, as passed by user.
- For some variables, rejects values that are clearly unrealistic (likely indicating improper units conversion or incorrect sign).

Additionally, CMOR facilitates compliance with CF through an IPCC-specific input table that provides

- Proper specification of several coordinate attributes, including:
  - → Correct standard name
  - → “axis”, “positive”, and “formula_terms” attributes, as appropriate
- Proper specification of several variable attributes, including:
  - → Correct standard name
  - → Required dimensions
  - → “cell_methods” attribute
- A capability to
  - → Reorder axis order
  - → Reverse axis direction (or translate longitude dimension)
  - → Convert units (through udunits)

How can we encourage CF compliance and "best practices" (i.e., inclusion of useful metadata)?

- Use CMOR to rewrite model output, relying on tables tailored to each specific project.

- These same tables can be used as input to a project-specific CF-checker to uncover missing or improperly stored metadata when CMOR is by-passed.

- Difficult to encourage "best practices" in general, but they may catch on if enforced in coordinated modeling activities.

Standard names are often insufficient to unambiguously describe fields.

Table A1a: **Monthly-mean 2-d atmosphere or land surface data (longitude, latitude, time:month).**

	CF standard_name	**output variable name**	**units**	**notes**
1	air_pressure_at_sea_level	psl	Pa	
8	surface_snow_thickness	snd	m	this thickness when multiplied by the average area of the grid cell covered by snow yields the time-mean snow volume. Thus, for time means, compute as the weighted sum of thickness (averaged over the snow-covered portion of the grid cell) divided by the sum of the weights, with the weights equal to the area covered by snow. report as 0.0 in snow-free regions.
15	surface_temperature	ts	K	"skin" temperature (i.e., SST for open ocean)
16	surface_air_pressure	ps	Pa	*not* mean sea-level pressure
19	atmosphere_water_vapor_content	prw	kg m^{-2}	vertically integrated through the atmospheric column
21	surface_runoff_flux	mrros	kg m^{-2} s^{-1}	compute as the total surface runoff leaving the land portion of the grid cell divided by the land area in the grid cell; report as "missing" or 0.0 where the land fraction is 0.
22	runoff_flux	mrro	kg m^{-2} s^{-1}	compute as the total runoff (including "drainage" through the base of the soil model) leaving the land portion of the grid cell divided by the land area in the grid cell; report as "missing" or 0.0 where the land fraction is 0.